Cardinality of Fuzzy Sets and Accumulation of Small Membership

Eduard Bartl and Radim Belohlavek, *Senior Member, IEEE*

***Abstract*—We describe an intuitive and practically significant empirical phenomenon that relates to the concept of cardinality of a fuzzy set, namely, an excessive accumulation of small degrees of membership. We argue and demonstrate by examples that the present notions of cardinality do not take this phenomenon into account properly and may thus prove insufficient in applications. We propose a new concept of cardinality, generalizing the well-known Zadeh's sigma count, demonstrate using both intuitive and technical examples that it alleviates the insufficiency of the existing ones, and provide a theoretical analysis of this concept. We also propose topics for future theoretical and empirical research.**

***Index Terms*—Cardinality, fuzzy set, sigma count.**

I. Sigma Count Cardinality of Fuzzy Sets

THE concept of cardinality of fuzzy sets has played an essential role since the early years of fuzzy sets. Among the various approaches to cardinality, arguably the most significant is the so-called scalar cardinality, and in particular, the so-called sigma count of a fuzzy set due to its intuitive appeal and practical relevance. Recall that for a fuzzy set A in a finite universe $U = \{u_1, \ldots, u_k\}$, i.e., $A : U \to [0, 1]$, the sigma count cardinality $|A|$ of A is defined by

$$|A| = A(u_1) + \cdots + A(u_k) \tag{1}$$

i.e., as the sum of the membership degrees $A(u_i)$. This concept was introduced under the term "power of fuzzy set" by De Luca and Termini [9]. Since the early 1970s, the concept has played a significant role in several of Zadeh's key papers on fuzzy sets; see, e.g., [3], [26, part III, p. 55], [28, p. 65], and [29, p. 31].[1] The sigma count eventually became a standard concept of (scalar) cardinality. It is widespread in textbooks on fuzzy logic as well as in research papers.[2]

Manuscript received 14 August 2023; revised 16 December 2023 and 18 February 2024; accepted 19 February 2024. Date of publication 29 March 2024; date of current version 4 June 2024. Recommended by Associate Editor D. Wu. *(Corresponding author: Eduard Bartl.)*

The authors are with the Department of Computer Science, Palacký University Olomouc, 779 00 Olomouc, Czech Republic (e-mail: eduard.bartl@upol.cz; radim.belohlavek@acm.org).

Digital Object Identifier 10.1109/TFUZZ.2024.3383279

[1]Interestingly, Zadeh [24, n. 3 on p. 7] defines the sigma count cardinality in his 1971 report, i.e., before De Luca and Termini [9], with no reference to De Luca and Termini. Since he attributes this concept to [9] in all of his subsequent writings, it is likely that he obtained this notion from De Luca and Termini before 1972, as he served as editor of *Information and Control* in which De Luca and Termini's paper appeared (the paper was received by the journal in July 1970).

[2]For a comprehensive treatment of the notion of cardinality of fuzzy sets, which includes both scalar and nonscalar cardinalities, we refer, e.g., to [16] and [23]; see also [5].

The aim of this article is to point out an intuitive phenomenon that reveals a significant drawback of the standard sigma count. We present examples demonstrating the phenomenon and propose a concept that alleviates the associated insufficiency of the standard sigma count. We articulate the limitations of the standard sigma count, which result from the intuitive phenomenon and which one must be aware of in applications of the cardinality of fuzzy sets.

II. Insufficiency of Sigma Count

Consider a city population U of 100 000 inhabitants and suppose 5000 of them have a university diploma. Suppose everyone in this population is rather slim, hence the concept "obese" applies to degree, say, 0.1 to each member of U. One may thus consider the crisp fuzzy set $Diploma$ of people who have a degree and a fuzzy set $Obese$ in which the degree of membership of each member of U equals 0.1.

Intuitively, there are more inhabitants with a diploma than those who are obese. Namely, the reasoning behind may be described as follows: "While there are 5000 persons with a degree, there is almost no one who actually may be considered obese. Hence the collection of people with a degree is larger than the collection of obese people."

However, using the sigma count cardinality (1), we obtain the opposite conclusion. Namely

$$|Diploma| = \underbrace{1 + \cdots + 1}_{\text{5000 times}} = 5000$$

and

$$|Obese| = \underbrace{0.1 + \cdots + 0.1}_{\text{100 000 times}} = 10\,000$$

whence

$$|Diploma| < |Obese|.$$

Such a conclusion contradicts intuition, and hence, presents a problem for the sigma count cardinality.

It is evident that the reason for obtaining this counterintuitive result is the accumulation of a large number of very small degrees of membership, which adds to the cardinality of the fuzzy set *Obese*. This becomes even more visible in a modified scenario with a larger population of inhabitants who are much slimmer.

We consider it important to demonstrate that the insufficiency of the sigma count cardinality not only becomes apparent in intuitive considerations, but also surfaces as a severe limitation in natural applications of cardinality. For this purpose, we present two examples demonstrating that employment of the sigma-count cardinality may lead to a wrong choice. While the first one is an illustrative example, the second one comes from a real-world study involving large-scale education data from the United Kingdom.

In our illustrative example, a municipality with 1000 residents needs to decide between two kinds of investment. The first one consists in building a water supply pipeline for a part of the town comprising 320 residents who still obtain water from local wells. The second investment consists in building a town cinema. To make a decision, a poll is organized in which each resident expresses his preference for each of the two options using a number from the interval $[0, 1]$. For the water pipeline option, each of the 320 residents assigns 0.75 (high preference yet smaller than 1 because water may presently be obtained from wells and is hence free of charge), while the remaining residents assign 0 (the remaining residents have no benefit from building the pipeline). For the cinema option, each resident expresses his preference by 0.25 (the residents have other options watching movies, such as subscription streaming services). This way, the two options are represented by the fuzzy sets P (pipeline) and C (cinema). Thus, e.g., $C(u) = 0.25$ for each respondent u. The municipality agrees on selecting the option preferred by most respondents. Now, since the number of residents preferring the water pipeline and the cinema are represented by $|P|$ and $|C|$, respectively, employment of the sigma-count cardinality results in selecting the cinema option because

$$|P| = 320 \cdot 0.75 = 240 < 250 = 1000 \cdot 0.25 = |C|.$$

Yet, such a choice is counterintuitive because while there is a substantial part of the residents who prefer the pipeline rather strongly, there is no resident in the town who substantially prefers building the cinema. Clearly, this the consequence of the accumulation of a large number of small degrees of preference for a cinema. As we demonstrate in Section III-C, the modified concept of cardinality developed in this article results in the right choice, i.e., the pipeline.

Our second example comes from data analysis of a large-scale education data gathered by the United Kingdom's governmental organizations. In fact, we realized the need to reconsider scalar cardinality when working on this project as the sigma count cardinality resulted in computing nonintuitive, flat factors [1], [2]. We now briefly describe the essence of the factor problem as it illustrates practical significance of the phenomenon involved. More on this problem can be found in the Appendix.

The problem relates to matrix distance, or dually, similarity of matrices. Consider two matrices, I and J, of dimension $n \times m$ with entries in the real unit interval $[0, 1]$. For instance, I and J may be regarded as representing two binary fuzzy relations between n elements (rows) and m fuzzy attributes (columns), i.e., the values I_{ij} and J_{ij} represent the truth degrees to which i and j are related by I and J, respectively. A natural way to measure the similarity $S(I, J)$ of I and J is via

$$S(I, J) = \sum_{i,j=1}^{n,m} (I_{ij} \leftrightarrow J_{ij}), \tag{2}$$

$$\begin{pmatrix}
0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 1.0 & 1.0 & 1.0 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 1.0 & 1.0 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 1.0 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 1.0 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 1.0 & 1.0 & 1.0 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6
\end{pmatrix}$$

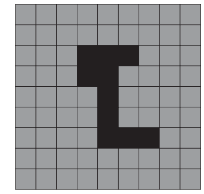

Fig. 1. Matrix I.

$$\begin{pmatrix}
0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 1.0 \\
0.2 & 0.2 & 1.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\
1.0 & 0.2 & 0.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\
1.0 & 0.2 & 0.2 & 0.0 & 0.0 & 0.2 & 0.2 & 0.2 & 0.2 \\
0.2 & 1.0 & 0.2 & 0.0 & 0.0 & 0.2 & 0.2 & 1.0 & 0.2 \\
0.2 & 0.2 & 0.2 & 1.0 & 0.2 & 0.2 & 1.0 & 0.0 & 0.2 \\
0.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.0 & 1.0 \\
0.0 & 0.2 & 0.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.0 & 1.0 \\
0.2 & 0.0 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 1.0 & 1.0
\end{pmatrix}$$

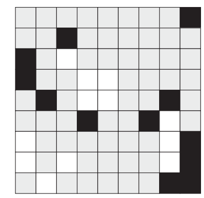

Fig. 2. Matrix J.

$$\begin{pmatrix}
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 1.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 1.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0
\end{pmatrix}$$

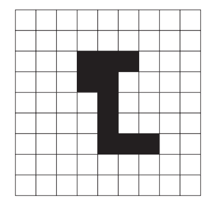

Fig. 3. Matrix K.

in which $I_{ij} \leftrightarrow J_{ij}$ is the biresiduum of I_{ij} and J_{ij}, and may hence be interpreted as a degree in $[0, 1]$ representing a proximity (similarity) of I_{ij} and J_{ij} [13], [14]. If $\leftrightarrow$ is, for instance, the Łukasiewicz equivalence, then $I_{ij} \leftrightarrow J_{ij} = 1 - |I_{ij} - J_{ij}|$ is a natural proximity. Formula (2) and its variations have been used in many situations [1], [4], [6], [7], [8], [10], [16]. Note that this formula is related to a generalization of the classical matrix L_1-norm (cf. Remark 1 below).

Consider now the universe set $U = \{1, \ldots, n\} \times \{1, \ldots, m\}$ and the fuzzy set $E : U \to [0, 1]$ representing a pair-by-pair similarity of I and J, which is defined by

$$E_{ij} = I_{ij} \leftrightarrow J_{ij}.$$

Clearly,

$$S(I, J) = |E|, \tag{3}$$

i.e., the matrix similarity $S(I, J)$ defined by (2) is the sigma count cardinality of the fuzzy set E associated to I and J.

To illustrate the problem of the sigma count, consider the matrices $I, J, K \in [0, 1]^{9\times 9}$ depicted in Figs. 1–3, along with their graphical representation (shades of gray represent membership degrees; the darker the shade, the higher the membership degree).

Consider the similarities $S(I, J)$ of I to J and $S(I, K)$ of I to K. It is a matter of simple calculation to verify that $S(I, J)$ is larger than $S(I, K)$, i.e., I is more similar to J than to K:

$$S(I, J) = \sum_{i,j=1}^{9,9} (I_{ij} \leftrightarrow J_{ij}) = 42.2,$$

$$S(I,K)=\sum_{i,j=1}^{9,9}(I_{ij}\leftrightarrow K_{ij})=38.4.$$

This result is a direct consequence of an accumulation of a large number of small similarity values $I_{ij}\leftrightarrow J_{ij}$, especially those with $I_{ij}=0.6$ (dark gray entries in I) and $J_{ij}=0.2$ or $J_{ij}=1$ (light gray or black entries in J), for which $I_{ij}\leftrightarrow J_{ij}=0.4$. An intuitive view, however, is different: I seems more similar to K than to J because there is a fair number of entries (those grouped in the L-shaped pattern) in which both I and K have exactly the same values.[3] As we demonstrate in Section III-C, this right conclusion is obtained when the modified concept of cardinality developed in this article is employed.

Remark 1: The relationship of matrix similarity to cardinality is not surprising in view of the well-known matrix metric that is based on the L_1-norm, which assigns to two real-valued matrices I and J the distance $d(I,J)=\sum_{i,j=1}^{n,m}|I_{ij}-J_{ij}|$. Namely, for binary matrices $I,J\in\{0,1\}^{n\times m}$, this $d(I,J)$ is the so-called Hamming distance (or simple matching distance), and equals the number of matrix entries for which I and J differ, i.e.,

$$d(I,J)=|\{\langle i,j\rangle\,;\,I_{ij}\neq J_{ij}\}|.$$

It is hence clear that $S(I,J)$ may be regarded as a dual (in the sense of similarity versus distance) to $d(I,J)$, since for binary matrices $S(I,J)=n\cdot m-d(I,J)$.

To sum up, the phenomenon of a possible accumulation of a large number of small membership degrees, which presents the problem in our intuitive example as well as the matrix similarity example, seems to point out a serious insufficiency of the sigma count cardinality. Needless to say, this phenomenon may impair applications of the sigma count cardinality in a variety of domains if ignored.

Yet, browsing the literature on cardinality of fuzzy sets, we found only one paper that explicitly discusses this phenomenon, namely by Ralescu [20]. In his interesting paper, Ralescu [20, p. 361 ff.] mentions—among several issues he examines—the undesirable effect of accumulation and claims that one of his propositions properly alleviates this effect. This is, however, not entirely the case as we show in the Appendix. In addition to [20], a comment can be found in Zadeh's first papers mentioning the sigma count. For example, in [28] and [29], Zadeh notes: "For some applications, it is necessary to eliminate from the count those elements of F whose grade of membership falls below a specified threshold. This is equivalent to replacing F in (4.70) with $F\cap\Gamma$, where Γ is a fuzzy or nonfuzzy set, which induces the desired threshold" (here, F is the fuzzy set of which the sigma count is computed).[4]

[3] This intuitive view of matrix similarity, in which a considerable number of highly similar entries are needed to make two matrices similar, is derived from factor analysis of data with fuzzy attributes; see Appendix. Clearly, for other applications, different notions of matrix similarity may be preferable.

[4] As is easily seen, using $F\cap\Gamma$ is, in fact, more general than eliminating the elements with membership below a given threshold.

III. New Approach to Cardinality of Fuzzy Sets

A. Definition

A natural way to alleviate the undesirable effect of accumulation of small degrees of membership by the standard sigma count (1) is to deemphasize the small degrees. A simple idea would be to disregard small degrees completely by selecting a threshold degree θ, such as $\theta=0.2$, in order to take into account only the elements with degree of membership not lower than θ, as proposed by Zadeh (see above). That is, to define the cardinality $|A|_\theta$ of a fuzzy set A by

$$|A|_\theta=\sum_{u\in U,A(u)\geq\theta}A(u). \tag{4}$$

However, this imposes an artificial sharp boundary: While the membership degree 0.2 is counted, $0.2-\epsilon$ is not, even though ϵ is arbitrarily small.

We hence propose to suppress small membership degrees, and also possibly emphasize large membership degrees, in a gradual manner using a modifying function c. We assume that the set L of truth degrees is the real unit interval $[0,1]_\mathbb{R}$ or, more generally, a subset of $[0,1]_\mathbb{R}$ with $0,1\in L$ (e.g., a finite subchain of $[0,1]_\mathbb{R}$, such as the five-element chain $\{0,0.25,0.5,0.75,1\}$).

Given such a scale L and its two elements $a\leq b$, we denote by $[a,b]$ the closed interval in L bounded by a and b, i.e.,

$$[a,b]=\{x\in L\mid a\leq x\leq b\}.$$

For this kind of scales L, we shall consider functions $c:L\to L$ modifying degrees of membership for which there exists $\theta\in(0,1]$ such that

$$c(a)\leq a \text{ for } a<\theta \text{ and}$$
$$c(a)\geq a \text{ for } a\geq\theta, \tag{5}$$
$$c \text{ is convex on } [0,\theta] \text{ and}$$
$$\text{concave on } [\theta,1]. \tag{6}$$

Remark 2: To apply appropriately to the general case of L, e.g., a finite chain, we consider the following formulation of convexity, which yields ordinary convexity for L being the real unit interval, in (6): c is convex on the interval $[d,e]\subseteq L$ if for every $a,b,x\in[d,e]$ with $a\leq x\leq b$ one has

$$c(x)\leq\left(1-\frac{x-a}{b-a}\right)c(a)+\frac{x-a}{b-a}c(b). \tag{7}$$

The meaning of this inequality is based on the following geometric view (see Fig. 4): for every $a,b,x\in[d,e]$, $a\leq x\leq b$, the value $c(x)$ lies below the value

$$s(x)=\left(1-\frac{x-a}{b-a}\right)c(a)+\frac{x-a}{b-a}c(b)$$

of the secant line of the function c connecting the points $\langle a,c(a)\rangle$ and $\langle b,c(b)\rangle$. The concept of concavity is approached analogously.

Note that if one wishes that only truth degrees from L appear in (7), one may replace the right-hand side by

$$\left\lfloor\left(1-\frac{x-a}{b-a}\right)c(a)+\frac{x-a}{b-a}c(b)\right\rfloor$$

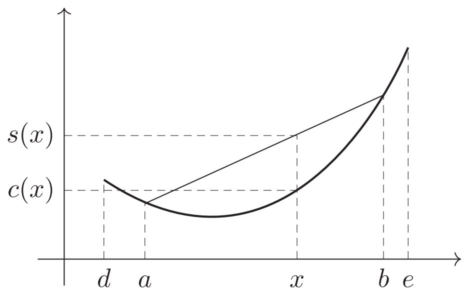

Fig. 4. Geometric interpretation of (7).

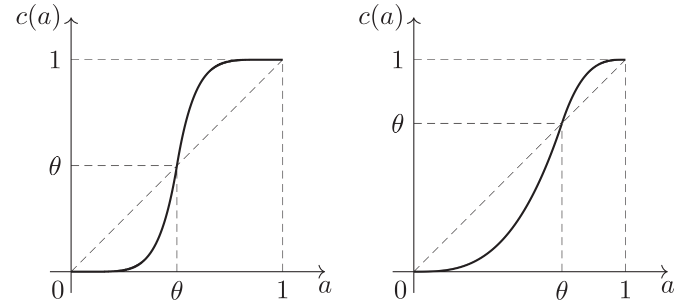

Fig. 5. Sigmoid-shape function (9). Left: $\theta = 0.5$ and $p = 6$; right: $\theta = 0.7$ and $p = 3$.

i.e., to round it down to the closest value that belongs to L. An inequality modified this way is equivalent with (7) and may be considered conceptually cleaner.

We now present a modification of the concept of sigma count cardinality, which alleviates the phenomenon addressed in Section II.

Definition 1: Let $c : L \to L$ be a function satisfying (5) and (6), and let $U = \{u_1, \ldots, u_k\}$. The cardinality $|A|_c$ of a fuzzy set $A : U \to L$ is defined by

$$|A|_c = c(A(u_1)) + \cdots + c(A(u_k)). \tag{8}$$

Remark 3: Let us comment on the meaning of conditions (5) and (6) imposed on the degree-modifying function c.

1) The conditions $c(a) \leq a$ (subdiagonality) and $c(a) \geq a$ (supradiagonality) express the basic requirement for c to suppress membership degrees smaller than the threshold θ and possibly emphasize those above the threshold.
2) The convexity on $[0, \theta]$ reflects the intuition regarding the intensity of suppression of membership degrees below the threshold: Around the threshold θ, a change in the membership degree $A(u)$ in general results in a more considerable difference in the perception of size of a given fuzzy set; as $A(u)$ moves toward 0, the same change in $A(u)$ results in a smaller difference in size of A.
3) A justification of the concavity requirement of c on $[\theta, 1]$ is symmetric to that of convexity in (2). Notice that concavity allows $c(a) = a$ for $a \geq \theta$, which—as we hypothesize—would be a possible intuitive choice.

Remark 4:

1) The idea to modify the membership degrees $A(u_i)$ in the sigma count formula (1) is not new. The authors of these approaches seem to have been motivated by formal reasons only, namely, by the fact that the modification provides a generalization of the basic sigma count.

   It appeared for the first time in Kaufmann's remarkable monograph [17, p. 43]. Kaufmann proposed to use the formula

$$|A|_p = [A(u_1)]^p + \cdots + [A(u_k)]^p$$

   for a positive integer p (in fact, he considered $p = 1$ and $p = 2$ only; the case of general p had then soon been adopted in [11]). Later on, a proposal to consider the modified sigma count using (8) was put forward in [22]; see also [23]. Wygralak [22] provides an axiomatic characterization of cardinalities and considers rather general functions c, namely those satisfying the boundary conditions $c(0) = 0$ and $c(1) = 1$ plus monotony. Our motivation leads to stronger requirements for c, and both the boundary conditions and monotony turn out to be consequences of our requirements; see Theorem 1.
2) Despite a thorough search of the literature, we have not found any considerations regarding applicational motivations for using Kaufmann's or the general Wygralak's modification of the basic sigma count.
3) As mentioned at the end of Section II, the undesirable accumulation of small membership degrees is discussed by Ralescu [20]. Ralescu's proposal to aleviate this effect is analyzed in the Appendix of our paper.
4) Very likely, the undesirable accumulation led Zadeh to his remark in [28] and [29], which we mention at the end of Section II. Observe that Zadeh's proposal to replace A by $A \cap \Gamma$ for a suitable fuzzy set Γ may be considered in the perspective of using a modifying function c. Namely, one may consider a modifying function $c(A(u)) = A(u) \wedge \Gamma(u)$, which, however, would be a concept of modification considerably more general than the one put forward in this article, because such a modification of $A(u)$ depends on u. That is, the consideration we describe would require a modifying function $c : U \times L \to L$ rather than just $c : L \to L$.

B. Examples of Modifying Function

We propose a sigmoid-shape function as a basic function that reflects our intuitive idea of suppressing small membership degrees and emphasizing large membership degrees. There is a variety of ways to define sigmoid-shape functions satisfying requirements (5) and (6). A straightforward one is based on the following concatenation of two complementary power functions:

$$c(a) = \begin{cases} \theta^{1-p} \cdot a^p, & \text{for } a < \theta \\ 1 - (1-\theta)^{1-p} \cdot (1-a)^p, & \text{for } a \geq \theta, \end{cases} \tag{9}$$

with $p \geq 1$ being a parameter determining the degree of nonlinearity of c; see Fig. 5. Clearly, θ is a fixpoint of c, i.e., $c(\theta) = \theta$. Note also that for $p > 1$ and $\theta = 0.5$, one obtains the

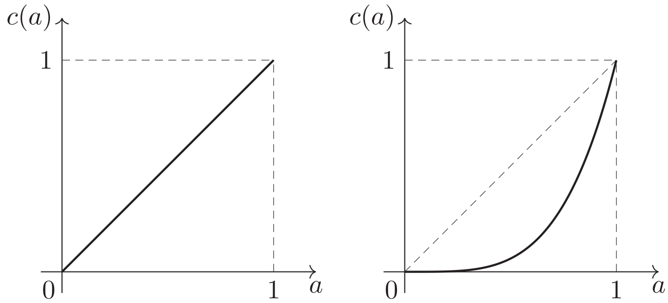

Fig. 6. Identity (left) and subdiagonal function (right).

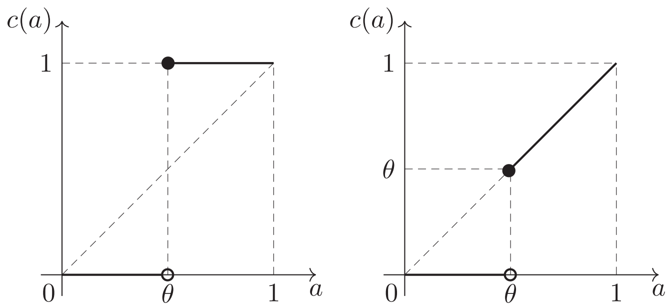

Fig. 7. Two variants of threshold functions.

contrast enhancement function proposed by Wygralak [22]; see Remark 4(1).[5]

Formula (9) defines a broad class of functions that may be used in various applications of cardinality. This class is convenient from a computational point of view because evaluating power functions on a computer is efficient.

Other examples of functions c satisfying conditions (5) and (6) are shown in Figs. 6 and 7.

1) The identity in Fig. 6 does not change membership degrees. Obviously, such function results in the ordinary sigma count cardinality, i.e., $|A|_c = \sum_{u\in U} A(u)$.
2) The subdiagonal function in Fig. 6 suppresses all degrees except 0 and 1, which remain unchanged. This function is an example of a truth-stressing linguistic hedge; see Section IV for further discussion.
3) The classic threshold function in the left part of Fig. 7 represents a simple way to modify membership degrees, albeit perhaps somewhat simplistic with respect to the aim to yield an appropriate concept of cardinality. Namely, the cardinality $|A|_c$ based on this threshold function yields

$$|A|_c = |{}^{\theta}\!/A| = |\{u \in U\,;\, A(u) \geq \theta\}|,$$

i.e., the ordinary cardinality (number of elements) of the θ-cut of A.
4) The right part of Fig. 7 depicts another variant of a threshold function, one for which the values greater than or equal to θ are not modified. This function yields formula (4) summing the membership degrees greater than or equal to θ.

[5]The importance of contrast enhancement functions for applications of fuzzy sets is mentioned already in [27].

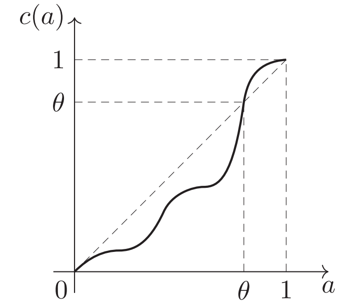

Fig. 8. Inappropriately selected function, which does not meet second part of (6).

Note that all of the aforementioned modifying functions are instances of a general class of functions c defined by

$$c(a) = \begin{cases} \theta^{1-p} \cdot a^p, & \text{for } a < \theta \\ 1-(1-\theta)^{1-r}\cdot(1-a)^r, & \text{for } a \geq \theta, \end{cases} \tag{10}$$

where $p \geq 1$ and $r \geq 1$ are parameters determining the degree of nonlinearity in the subdiagonal and supradiagonal parts of c, respectively. In particular, we obtain the sigmoid-shape function (9) for $p = r$; the identity function for $p = r = 1$; the subdiagonal function for $p > 1$ and $\theta = 1$; the threshold function in the left part of Fig. 7 for $p, r \to \infty$; and the threshold function in the right part of Fig. 7 for $p \to \infty$ and $r = 1$.

Fig. 8 presents a function c that suppresses small membership degrees and emphasizes large ones but does not meet the intuitive requirements discussed previously. Namely, the rates of modification of membership degrees below the threshold θ are arbitrary and do not conform to the requirement of convexity on $[0, \theta)$. Note that this function still satisfies all conditions presented in [23, Th. 3.2].

C. Solutions to the Problems in Section II

Consider again the examples in Section II demonstrating the inadequacy of the sigma count cardinality. The proposed modified concept of cardinality naturally solves the problems in these examples. For the first example, consider, for instance, the sigmoid-shape function (9) for c with $\theta = 0.5$ and $p = 4$. One then easily verifies that

$$|Diploma|_c = \underbrace{1 + \cdots + 1}_{\text{5000 times}} = 5000,$$

and

$$|Obese|_c = \underbrace{0.008 + \cdots + 0.008}_{\text{100 000 times}} = 80.$$

As a result, one obtains

$$|Diploma|_c \gg |Obese|_c,$$

which meets the intuitive requirement.

In the municipal voting example, consider the sigmoid-shape function c with $\theta = 0.5$ and $p = 2$. For such a function, it holds $c(0.25) = 0.125$ and $c(0.75) = 0.875$, therefore

$$|P|_c = 320 \cdot 0.875 = 280 > 125 = 1000 \cdot 0.125 = |C|_c,$$

which again agrees with the intuition.

For the matrix similarity example, consider again the sigmoid-shape function (9), now with $\theta = 0.75$ and $p = 9$. Due to the relatively large p, the modifying function c is steeply increasing around θ, which results in a significant suppression of the membership degrees that lead to the undesired accumulation with the standard sigma count. As a result, one obtains

$$S_c(I,J) = \sum_{i,j=1}^{9,9} c(I_{ij} \leftrightarrow J_{ij}) = 6.3, \quad \text{and}$$

$$S_c(I,K) = \sum_{i,j=1}^{9,9} c(I_{ij} \leftrightarrow K_{ij}) = 10.2,$$

i.e., I and K appear more similar than I and J when using our modified concept of cardinality.

D. Properties of the Modifying Function

As the following theorem shows, our requirements of the modifying functions c entail two elementary conditions for c, the boundary conditions and monotony. The boundary condition $c(0) = 0$ expresses that the membership degree 0 does not contribute to cardinality at all, while $c(1) = 1$ implies that 1 contributes to the largest extent possible. The monotony condition expresses the basic requirement that larger membership degrees contribute more than smaller ones. Both the entailed conditions may hence be regarded as necessary ones that must be satisfied by every c for the modified sigma count to represent a reasonable approach to cardinality at all.

Theorem 1: Let $c : L \to L$ satisfy (5) and (6). Then

1) c satisfies the boundary conditions

$$c(0) = 0 \text{ and } c(1) = 1 \tag{11}$$

2) c is monotone, i.e.,

$$a \leq b \text{ implies } c(a) \leq c(b).$$

Proof:

1) Since $0 < \theta$, the condition $c(0) = 0$ follows from the subdiagonality, $c(a) \leq a$, for $a < \theta$, while $c(1) = 1$ follows from the supradiagonality of c.
2) Let us check that the monotony of c on $[0, \theta]$ follows from the convexity of c on $[0, \theta]$. Suppose, by contradiction, that c not monotone, i.e., $c(b_1) > c(b_2)$ for some $0 \leq b_1 < b_2 \leq \theta$. Since $\frac{b_1}{b_2} c(b_2)$ is the vertical coordinate of the point with the horizontal coordinate b_1 on the line connecting the points $\langle 0, 0\rangle$ and $\langle b_2, c(b_2)\rangle$, we clearly have $0 \leq \frac{b_1}{b_2} c(b_2) \leq c(b_2) < c(b_1)$, which contradicts the convexity of c on $[0, \theta]$, since for $a = 0$, $x = b_1$, and $b = b_2$, (7) yields

$$c(b_1) \leq \left(1 - \frac{b_1 - 0}{b_2 - 0}\right) c(0) + \frac{b_1 - 0}{b_2 - 0} c(b_2) = \frac{b_1}{b_2} c(b_2).$$

In a similar manner, one proves that the monotony of c on $[\theta, 1]$ follows from the concavity of c on $[\theta, 1]$. □

A natural condition for the threshold θ is arguably to be a fixpoint of c, i.e., $c(\theta) = \theta$, which means that the membership degree θ contributes by θ to the cardinality of the given fuzzy set. If this is the case, the subdiagonality and supradiagonality conditions appear redundant, and may hence be omitted from our requirements on c.

Theorem 2: If θ is a fixpoint of c, i.e., $c(\theta) = \theta$, then the subdiagonality and supradiagonality conditions (5) follow from the convexity and concavity conditions (6).

Proof: The convexity of c on $[0, \theta]$, i.e., (7) for $a = 0$ and $b = \theta$, yields

$$c(x) \leq \left(1 - \frac{x - 0}{\theta - 0}\right) c(0) + \frac{x - 0}{\theta - 0} c(\theta) = \frac{x}{\theta}\theta = x,$$

for every $x \in [0, \theta]$, i.e., the subdiagonality of c on $[0, \theta]$.

Analogously, the concavity of c on $[\theta, 1]$, i.e., condition (7) with $\geq$ replacing $\leq$ and with $a = \theta$ and $b = 1$, yields

$$\begin{aligned} c(x) &\geq \left(1 - \frac{x - \theta}{1 - \theta}\right) c(\theta) + \frac{x - \theta}{1 - \theta} c(1) \\ &= \left(1 - \frac{x - \theta}{1 - \theta}\right) \theta + \frac{x - \theta}{1 - \theta} 1 \\ &= \theta + \frac{x - \theta}{1 - \theta}(1 - \theta) = \theta + x - \theta = x, \end{aligned}$$

for every $[\theta, 1]$, i.e., the supradiagonality of c. □

We now provide an axiomatic characterization of cardinalities defined by (8) in the style of Wygralak [22]. Below, $\{^a/u\}$ denotes the fuzzy set A for which $A(v) = a$ for $v = u$ and $A(v) = 0$ for $v \neq u$.

Theorem 3: Let U be a finite universe and $L \subseteq [0, 1]$ contain 0 and 1. A function $\text{card} : L^U \to [0, |U|]$ satisfies

$$\text{card}(A) = |A|_c$$

for every fuzzy set $A : U \to L$, where $|A|_c$ is defined by (8) for some function $c : L \to L$ satisfying (5) and (6), if and only if card satisfies the following conditions:

1) For each $u \in U$:

$$\text{card}\left\{^1/u\right\} = 1.$$

2) There exists $\theta > 0$ in L such that

2.1) for each $u \in U$;

$$\begin{aligned} &\text{card}\{^a/u\} \leq a \text{ for } a < \theta \quad \text{and} \\ &\text{card}\{^a/u\} \geq a \text{ for } a \geq \theta; \end{aligned}$$

2.2) for each $u, v \in U$ and every $0 \leq a \leq x \leq b \leq \theta$

$$\begin{aligned} \text{card}\{^x/u\} \leq &\left(1 - \frac{x - a}{b - a}\right) \text{card}\{^a/v\} \\ &+ \frac{x - a}{b - a} \text{card}\left\{^b/v\right\}, \end{aligned} \tag{12}$$

and for each $u, v \in U$ and every $\theta \leq a \leq x \leq b \leq 1$

$$\begin{aligned} \text{card}\{^x/u\} \geq &\left(1 - \frac{x - a}{b - a}\right) \text{card}\{^a/v\}, \\ &+ \frac{x - a}{b - a} \text{card}\left\{^b/v\right\}. \end{aligned} \tag{13}$$

3) For any $n \geq 1$ and a collection of mutually disjoint $A_1, \ldots, A_n \in L^U$, i.e., $\min(A_i(u), A_j(u)) = 0$ for $i \neq j$ and each $u \in U$, one has

$$\mathrm{card}(A_1 \cup \cdots \cup A_n) = \mathrm{card}(A_1) + \cdots + \mathrm{card}(A_n).$$

Proof: If c satisfies (5) and (6), then $|\cdot|_c$ defined by (8) satisfies the conditions 1)–3) of the theorem. Indeed, 1) follows from (11) in Theorem 1, and 2.1) and 2.2) are direct consequences of (5) and (6), respectively. As for 3), the assumption of disjointness implies that for each $u \in U$ there is at most one i such that $A_i(u) > 0$. Hence if $(A_1 \cup \cdots \cup A_n)(u) > 0$, then there is a unique i such that

$$(A_1 \cup \cdots \cup A_n)(u) = A_i(u) > 0.$$

This property along with $c(0) = 0$, which holds due to Theorem 1, now easily entails 3).

Conversely, we prove that for the function $c : L \to L$ defined by

$$c(a) = \mathrm{card}\{{}^{a}/u\} \tag{14}$$

for any $u \in U$, one obtains $\mathrm{card}(A) = |A|_c$ for each $A \in L^U$. We first show that $c(a)$ is defined correctly, i.e., that $c(a)$ does not depend on u and that $c(a) \in L$ for each $a \in L$.

Observe first that since $0 < \theta \leq 1$, 2.1) implies $\mathrm{card}\{{}^{0}/u\} = 0$ and $\mathrm{card}\{{}^{1}/u\} = 1$ for each $u \in U$. Next, we show that

$$\mathrm{card}\{{}^{a}/u\} = \mathrm{card}\{{}^{a}/v\} \tag{15}$$

for every $u, v \in U$. Let first $0 \leq b_1 \leq b_2 \leq \theta$. Putting $a = 0$, $x = b_1$, and $b = b_2$, inequality (12) yields

$$\begin{aligned}\mathrm{card}\{{}^{b_1}/u\} &\leq \left(1 - \frac{b_1 - 0}{b_2 - 0}\right) \mathrm{card}\left\{{}^{0}/v\right\} \\ &\quad + \frac{b_1 - 0}{b_2 - 0} \mathrm{card}\left\{{}^{b_2}/v\right\} \\ &= \left(1 - \frac{b_1}{b_2}\right) \cdot 0 + \frac{b_1}{b_2} \mathrm{card}\left\{{}^{b_2}/v\right\} \\ &= \frac{b_1}{b_2} \mathrm{card}\left\{{}^{b_2}/v\right\} \leq \mathrm{card}\left\{{}^{b_2}/v\right\}.\end{aligned}$$

Since $a \leq a$ for each $a \in L$, the inequality we just proved implies that $\mathrm{card}\{{}^{a}/u\} \leq \mathrm{card}\{{}^{a}/v\}$. Analogously, from inequality (13), we obtain $\mathrm{card}\{{}^{a}/u\} \geq \mathrm{card}\{{}^{a}/v\}$, verifying (15).

Next, in view of (15), $c(a)$ in (14) does not depend on u. Since $a \leq 1$ for each $a \in L$, we obtain

$$c(a) = \mathrm{card}\{{}^{a}/u\} \leq \mathrm{card}\{{}^{1}/u\} = 1,$$

proving that $c(a) \in L$ for each $a \in L$. Put together, (14) indeed provides a correct definition of a function $c : L \to L$.

Consider now an arbitrary fuzzy set $A \in L^U$ and let $U = \{u_1, \ldots, u_k\}$. Since

$$A = \left\{{}^{A(u_1)}/u_1\right\} \cup \cdots \cup \left\{{}^{A(u_k)}/u_k\right\},$$

3) implies

$$\begin{aligned}\mathrm{card}(A) &= \mathrm{card}\left\{{}^{A(u_1)}/u_1\right\} + \cdots + \mathrm{card}\left\{{}^{A(u_k)}/u_k\right\} \\ &= c(A(u_1)) + \cdots + c(A(u_k)) \\ &= |A|_c,\end{aligned}$$

finishing the proof. □

Remark 5: Alternatively, the following variants of conditions 1)–3) of Theorem 3 may be used.

1) Instead of 1), one may use the following requirement: For each crisp fuzzy set A in U, the value $\mathrm{card}(A)$ equals the classical cardinality of the ordinary set corresponding to A, i.e., $\mathrm{card}(A)$ is the number of $u \in U$ for which $A(u) = 1$. It is easy to see that due to 3), condition 1) is equivalent to the new requirement.
2) Instead of 2.2), one may use the corresponding weaker inequalities resulting from those in 2.2) for $u = v$, along with a new condition requiring that for each $u, v \in U$ and $a \in L$, one has $\mathrm{card}\{{}^{a}/u\} = \mathrm{card}\{{}^{a}/v\}$.

IV. Conclusion and Future Research

This article aims to point out a significant drawback of the standard sigma-count cardinality of fuzzy sets and propose a modification that alleviates this drawback. The drawback is substantial from an applicational viewpoint, which we demonstrate by intuitive examples and examples from factor analysis of data with fuzzy attributes. The essence of the drawback consists in a possible accumulation of small membership degrees, making the size of a given fuzzy set large even though the fuzzy set appears intuitively small. As a result, employment of the standard sigma-count cardinality may lead to wrong conclusions if decisions involve the size of fuzzy sets.

We propose a new concept of cardinality that naturally suppresses the contribution of small membership degrees via a modifying function whose properties reflect intuitive requirements regarding the size of a fuzzy set. We demonstrate using examples that the thus modified concept of cardinality alleviates the drawback of the classic sigma-count cardinality and leads to correct results. Furthermore, we propose examples and a parameterized family of the modifying functions, and study the properties of the modified concept.

Even though the concept of sigma-count cardinality is one of the textbook concepts in the fuzzy set theory, the undesirable effect of accumulating small membership degrees and its practical ramifications are virtually not discussed in the literature. An exception to this situation is Ralescu's paper [21] that presents a solution different from our proposition that we critically examine.

While this article focuses on identifying the drawback of the standard sigma count and our modified concept of cardinality, along with illustrative examples and basic theoretical analysis, several questions remain for future exploration. These include the following.

1) *Selection of a concrete modifying function:* Our conditions for the modifying functions c, presented in Section III-A, reflect the intuition regarding a proper measurement of the size of fuzzy sets and delineate a large class of reasonable functions; cf. Section III-B. In a particular application of cardinality, however, one needs to select a single function c from this class. That is, to select a function c representing an intuitively reasonable concept of cardinality $|\cdot|_c$. The

difficulty of this task consists in that being intuitively reasonable is subjective and may also depend on the context.[6]
Apart from certain situations, in which the cardinality $|\cdot|_c$ is employed in a problem to which a certain criterion is associated, and hence, the choice of c may correspond to maximization or minimization of the criterion, selection of c needs to be based on intuitively sound and possibly also psychologically justified rules. Exploration of such rules presents an important practical problem. The problem, though, is of a foundational nature because its essence is directly related to the fundamental question of *how do people assess the size of a collection with a graded membership?*
Given our approach to cardinality, it seems natural to base the choice of c on questions of the following kind: How many elements with membership degree 0.1 needs a fuzzy set contain in order to have the size equal to 1, i.e., the size of a singleton. That is, what is the k for which

$$\left|\{{}^{0.1}/u_1,\ldots,{}^{0.1}/u_k\}\right|_c = \left|\{{}^{1}/u\}\right|_c ?$$

One may explore similar questions for other truth degrees, e.g., what k satisfies $|\{{}^{a}/u_1,\ldots,{}^{a}/u_k\}|_c = |\{{}^{1}/u\}|_c$ for other $a > 0$? Such questions seem to be amenable to human judgment and provide concrete information about the values of c. Namely, due to the definition of $|\cdot|_c$, the number k clearly implies $c(a) = \frac{1}{k}$.

2) *Modifying functions as linguistic hedges:* The modifying functions c we employ in our concept shall be examined in the context of linguistic hedges. In fuzzy logic, linguistic hedges are understood as expressions "very" or "roughly," which modify the meaning of fuzzy sets representing natural language expressions such as "very cold" [18], [25].
From a logical viewpoint, hedges are commonly regarded as unary connectives interpreted by functions $c : L \to L$ transforming truth degrees to (modified) truth degrees satisfying appropriate conditions, usually including $c(0) = 0$ and $c(1) = 1$. Two basic types of hedges are distinguished: Truth-stressing (or intensifying), such as "very," and truth-depressing (or relaxing), such as "roughly." The functions c interpreting these two types of hedges are required to be subdiagonal and supradiagonal, i.e., satisfy $c(a) \le a$ and $c(a) \ge a$, respectively, for each $a \in L$.
From this viewpoint, the functions c interpreting the truth-stressing hedges as well as those interpreting the truth-depressing hedges are particular instances of the modifying functions we consider; see Section III-B and Fig. 6 (right) for the truth-stressing case. Since our proposal's basic idea is to modify the membership degrees by an appropriate function c, it is only natural to regard our modifying functions as linguistic hedges and consider them within this context.
The nature of such rather general hedges from the logical and linguistic viewpoints then needs to be explored, as well as their further algebraic and logical properties. For one, an axiomatization of such hedges seems to present a nontrivial problem; see, e.g., [15] and [21] for axiomatizations of particular classes of truth-stressing and truth-depressing hedges, and [12] for an approach covering a considerably more extensive class of these two types of hedges.
Second, we contend that our modifying functions might provide a more proper meaning certain hedges, such as "more or less," compared to the commonly employed truth-depressing functions. Therefore, considerations regarding this kind of hedge should be looked at from the presented perspective.

3) *The very concept of cardinality of fuzzy sets:* Our proposal of $|\cdot|_c$ is to be regarded as a straightforward way to alleviate the drawback of the standard sigma-count cardinality. Other possible approaches should result from further exploring the question of how people assess the size of collections with a graded membership. In this regard, several ideas seem worth to be considered. One, implicitly present in Zadeh's early note (cf. Remark 4(4)), is to consider modifying functions c dependent on the elements $u \in U$, i.e., to consider them as functions

$$c : U \times L \to L$$

rather than $c : L \to L$. Another idea, which actually proved useful in our experiments with factor analysis of data with fuzzy attributes [1], [2], is to consider c dependent on the number of elements in the universe U. Intuitively, the larger the universe, the more suppression by c of the membership degrees is needed to obtain an intuitive notion of size. Last but not least, the phenomenon of undesirable accumulation of small membership degrees shall be taken into account in studies of the so-called nonscalar cardinalities of fuzzy sets.

4) *Psychological considerations:* During our work on the present topic, we talked to a number of people with varying degrees of mathematical training. We found that particularly people with no experience with the concept of a fuzzy set had difficulties understanding questions about the size of fuzzy sets, such as "How many of the thirty people in the class are obese?" in particular if the membership degrees in the fuzzy set are low. In this particular case, it came out of the ensuing discussions that perhaps a different term, such as "size (measure) of obesity in the class," could turn out as more appropriate than "number of obese people."
In general, the discussions revealed that the concept of a size of a vaguely delineated collection of objects is not as straightforward as it might seem. As a consequence, the concept itself demands a careful psychological exploration. While psychology, and in particular, the psychology of concepts, includes substantial work on human categories in which membership is a matter of degree [19], the concept of size has not been explored yet to the best of our knowledge. In addition to help us understand how humans perceive the size of vaguely delineated groupings, such explorations are also likely to put on firmer ground the various intuitive requirements on which the properties

[6]In principle, this task is similar to the need to select particular membership functions of fuzzy sets representing the meaning of linguistic terms such as "high temperature," which occurs in many applications of fuzzy sets.

of the concept of cardinality of fuzzy sets are based; cf. 1) in this section.

5) *Reconsideration of past studies and applications of cardinality:* The observations put forward in this article ask for a reconsideration of some previous studies involving the concept of cardinality of fuzzy sets. For one, it seems logical to revisit previous applications of the standard sigma-count cardinality from the present perspective. As mentioned in Section II, our motivation to propose a new concept of cardinality came from difficulties we encountered when applying the standard sigma-count cardinality. Employment of the modified concept delivered considerably better results and a similar improvement is to be expected in other applications involving the standard cardinality.

In addition, since the concept of cardinality of fuzzy sets explicitly or implicitly appears in several notions of the fuzzy set theory that are frequently used in applications, reconsidering these notions also appears as a natural step. As an example, various notions of similarity of fuzzy sets are based on the sigma-count cardinality [8], such as the basic one defined for fuzzy sets $A, B : U \to [0,1]$ by

$$\mathrm{sim}(A,B) = \sum_{u \in U} 1 - |A(u) - B(u)|$$

or its normalized version

$$\frac{\sum_{u \in U} 1 - |A(u) - B(u)|}{|U|}.$$

With the Łukasiewicz logical connectives, the similarity of matrices used previously, cf. (2), derives as a particular case of this basic similarity. It is apparent that $\mathrm{sim}(A,B)$ is but the standard sigma-count cardinality $|E|$ of the fuzzy set $E(u) = 1 - |A(u) - B(u)|$. As such, this concept may be regarded as inadequate due to the effect of accumulation of a small degrees. In particular, it may yield a reasonably high value of similarity to fuzzy sets A and B for which $A(u)$ is rather different from $B(u)$ for each $u \in U$.

Appendix A
Avoiding Flat Factors Via the Proposed Concept of Cardinality

We now illustrate the undesired effect of the standard sigma count using an example in factor analysis of data with fuzzy attributes, which actually led us to realize the problem discussed in this article and which we mentioned in Section II. The example comes from a large-scale factor analysis of educational data gathered by the United Kingdom's governmental organizations [1].[7]

Put briefly, the essence of factor analysis of an $n \times m$ data matrix I, for which each entry I_{ij} is interpreted as the truth degree to which the object i has the attribute j, is to find a small number k of factors explaining the data represented by I.

[7]The factor analysis study was initiated by a leading expert in analyzing educational data, Alex Scharaschkin. For concepts and technical details involved as well as for further references regarding factor analysis we refer to [1].

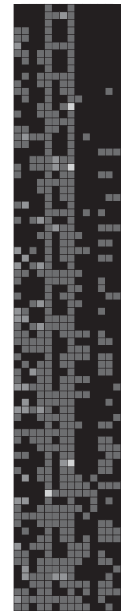

Fig. 9 First 80 rows of input data matrix I.

This is accomplished by computing a decomposition $I \approx A \circ B$ into an $n \times k$ object-factor matrix A and a $k \times m$ factor-attribute matrix B. In this decomposition, the lth factor ($l = 1, \ldots, k$) is represented by the lth column $A_{\_l}$ of A and the lth row $B_{l\_}$ of B. The cross products $F_l = A_{\_l} \circ B_{l\_}$, called factors in the following, are matrices that enable to reconstruct the input matrix I. In particular, the biresiduum $I_{ij} \leftrightarrow (F_l)_{ij}$ represents the degree to which the relationship between the object i and the attribute j is explained by the factor F_l. The factorization algorithms compute the factors in the decomposition one by one according to their explanatory power, which is defined as the sum $S(I, F_l)$, defined by (2), of the values $I_{ij} \leftrightarrow (F_l)_{ij}$, i.e., according to (3) as the sigma-count cardinality.

Now, as we observed in [2] in factor analysis study of educational data involving extensive examination tests used in the United Kingdom, using this approach may produce "flat" factors that do not have the desired ability to explain the input data. Namely, it may happen that of two possible factors, say F and G, the factor F is selected because $S(I,F) > S(I,G)$ even though intuitively, G is regarded as having a considerably higher explanatory power. This often occurs due to the undesirable accumulation of small degrees $I_{ij} \leftrightarrow (F)_{ij}$, which results in $S(I,F) > S(I,G)$. Since $S(\cdot,\cdot)$ is in fact the sigma-count cardinality, cf. (3), the selection of the undesirable factor is a consequence of using the sigma-count cardinality.

To see a concrete example, consider Fig. 9, which displays the first 80 rows of a 607×14 matrix I representing 607 best students (with a total mark A) and their 14 marks on a five-element scale L, evaluating the students' performance in six subject areas with regard to certain assessment objectives. The GreConD factorization algorithm produces several factors for this matrix I, two of which, F and G, are depicted in Fig. 10.

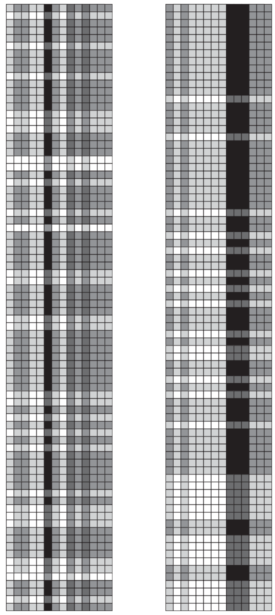

Fig. 10 First 80 rows of factors F (left) and G (right).

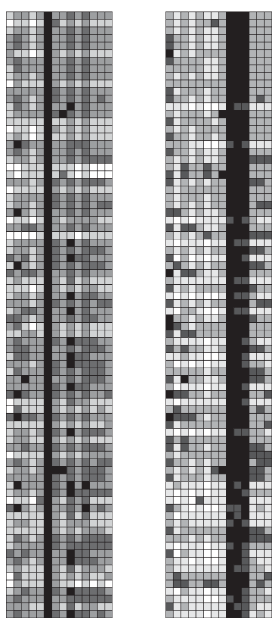

Fig. 11 First 80 rows of the biresidua matrices $I \leftrightarrow F$ (left) and $I \leftrightarrow F$ (right).

The corresponding matrices of biresidua, $I \leftrightarrow F$ and $I \leftrightarrow G$, are shown in Fig. 11. The algorithm prefers F over G since

$$S(I, F) = |E_F| = 4400 > 4372 = |E_G| = S(I, G),$$

where $|E_F|$ and $|E_G|$ are the sigma-count cardinalities of the fuzzy sets defined by $(E_F)_{ij} = I_{ij} \leftrightarrow F_{ij}$ and $(E_G)_{ij} = I_{ij} \leftrightarrow G_{ij}$; cf. (3). However, the biresidua matrices in Fig. 11 clearly reveal that G has intuitively a considerably higher explanatory power. Namely, most entries of the input matrix I are explained by F to a lower extent only, i.e., quite poorly, since the entries of $I \leftrightarrow F$ are mostly gray. Only a few entries of I are explained to a high degree by F. On the other hand, G explains a considerable part of I to a high degree (the dark rectangular area of $I \leftrightarrow G$). In this sense, while G is good, focused factor, F is flat.

Replacing the standard sigma-count $|\cdot|$ by our modified cardinality $|\cdot|_c$, in which c is defined by (10) with $\theta = 0.75$, $p = 5$, and $q = 1$, resolves the problem because one then obtains

$$S_c(I, F) = |E_F|_c = 2563 < 2651 = |E_G|_c = S_c(I, G).$$

That is, with the modified concept of cardinality, a proper factor gets selected by the algorithm.

Appendix B
On Ralescu's Paper "Cardinality, Quantifiers, and the Aggregation of Fuzzy Criteria"

The insufficiency of the sigma count is mentioned by Ralescu [20]. In particular, Ralescu argues that the sigma count "may be a relatively large number due to the cumulative effect of adding a large number of small quantities. In those cases σ-count A gives a counter-intuitive answer to the question 'how many elements are in A'?" Although this approach to cardinalities of fuzzy sets resolves some problems with the sigma count, it does not do so in a fully satisfactory manner, as we demonstrate below.

Recall first the concept of cardinality of fuzzy set presented in [20]. Ralescu defines a fuzzy cardinality of a fuzzy set $A : U \to L, U = \{u_1, \ldots, u_n\}$ as the fuzzy set $|A|_{\mathrm{Rf}}$ assigning to a nonnegative integer $k = 0, 1, \ldots, n$, the degree

$$|A|_{\mathrm{Rf}}(k) = \min \left\{ a_{(k)}, 1 - a_{(k+1)} \right\},$$

where $a_{(1)}, \ldots, a_{(n)}$ are the membership degrees $A(u_1), \ldots, A(u_n)$ ordered in a nonincreasing manner, and $a_{(0)} = 1$, $a_{(n+1)} = 0$. The scalar cardinality $|A|_{\mathrm{Rs}}$ of a fuzzy set A is then defined as follows:

$$|A|_{\mathrm{Rs}} = \begin{cases} 0, & \text{if} \quad A = \emptyset, \\ j, & \text{if} \quad a_{(j)} \geq 0.5, \\ j-1, & \text{if} \quad a_{(j)} < 0.5, \end{cases}$$

where $j = \max\{1 \leq k \leq n \,|\, a_{(k-1)} + a_{(k)} > 1\}$. Ralescu then claims that $|A|_{\mathrm{Rs}}$ equals one of the integer k at which $|A|_{\mathrm{Rf}}$ reaches its maximum, i.e.,

$$|A|_{\mathrm{Rf}}(|A|_{\mathrm{Rs}}) = \max_{0 \leq k \leq n} \{|A|_{\mathrm{Rf}}(k)\}.$$

On the one hand, Ralescu's approach leads to the intuitively correct conclusion as regards the cardinality of the fuzzy sets *Obese* and *Diploma* presented in Section II. Indeed, $|Diploma|_{\mathrm{Rs}} = 5000$, $|Obese|_{\mathrm{Rs}} = 0$, hence $|Diploma|_{\mathrm{Rs}} \gg |Obese|_{\mathrm{Rs}}$.

On the other hand, consider the fuzzy set A with the degree of membership of each element of U equal to, say, 0.49. Intuitively, the cardinality of this fuzzy set is larger than the cardinality of *Obese*, as

$$A = \left\{ {}^{0.49}/u_1, \ldots, {}^{0.49}/u_{100\,000} \right\} \text{ and}$$

$$Obese = \left\{ {}^{0.1}/u_1, \ldots, {}^{0.1}/u_{100\,000} \right\}.$$

Contrary to this intuition, Ralescu's approach renders

$$|A|_{\mathrm{Rs}} = |Obese|_{\mathrm{Rs}} = 0.$$

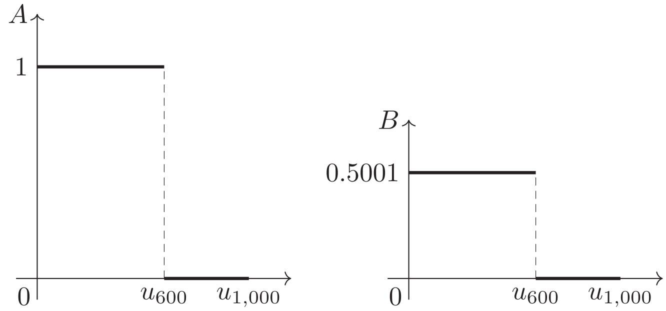

Fig. 12 Fuzzy sets A and B with cardinalities $|A|_{\mathrm{Rs}} = 600$ and $|B|_{\mathrm{Rs}} = 600$.

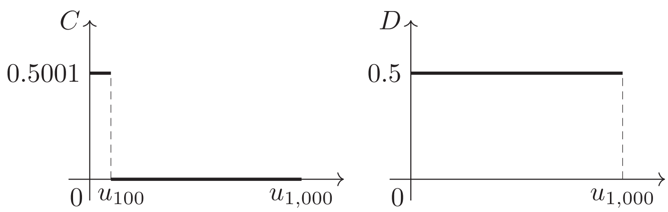

Fig. 13 Fuzzy sets C and D with cardinalities $|C|_{\mathrm{Rs}} = 100$ and $|D|_{\mathrm{Rs}} = 1$.

To make the situation more explicit, consider the fuzzy sets $A, B, C, D : U \to [0, 1]$, $U = \{u_1, \ldots, u_{1000}\}$, defined by

$$A(u_i) = \begin{cases} 1, & \text{for } i\text{=1,…,600}, \\ 0, & \text{otherwise}, \end{cases}$$

$$B(u_i) = \begin{cases} 0.5001, & \text{for } i\text{=1,…,600}, \\ 0, & \text{otherwise}, \end{cases}$$

$$C(u_i) = \begin{cases} 0.5001, & \text{for } i\text{=1,…,100}, \\ 0, & \text{otherwise}, \end{cases}$$

$$D(u_i) = 0.5, \text{ for } i = 1, \ldots, 1000.$$

These fuzzy sets are depicted in Figs. 12 and 13. Clearly, the cardinality of A should be larger than the cardinality of B, and the cardinality of C should be noticeably smaller than the cardinality of D. However, based on Ralescu's approach, we obtain $|A|_{\mathrm{Rs}} = 600$, $|B|_{\mathrm{Rs}} = 600$, $|C|_{\mathrm{Rs}} = 100$, and $|D|_{\mathrm{Rs}} = 1$, contradicting the intuitive expectations.

To conclude, Ralescu's concept of cardinality is able to alleviate the cumulative effect of small membership degrees. However, it does not distinguish properly between fuzzy sets belonging to certain classes as follows:

1) fuzzy sets with all nonzero degrees greater than 0.5 have the cardinality equal to the number of these degrees;
2) fuzzy sets with all degrees smaller than or equal to 0.5 have cardinality 0 or 1, irrespective of the size of the universe U.

As a result, Ralescu's approach does not treat the degrees of membership in accordance with intuition. As is easily seen, our approach to cardinality does not suffer from this problem.

References

[1] E. Bartl and R. Belohlavek, "Avoiding flatness in factoring ordinal data," *Inf. Sci.*, vol. 629, pp. 471–487, 2023.
[2] E. Bartl, R. Belohlavek, and A. Scharaschkin, "Toward factor analysis of educational data," in *Proc. CLA*, 2018, pp. 191–206.
[3] R. E. Bellman and L. A. Zadeh, "Local and fuzzy logics," in *Modern Uses of Multiple-Valued Logic*, J. M. Dunn and G. Epstein Eds. Dordrecht, Netherlands: Springer, 1977.
[4] R. Belohlavek, *Fuzzy Relational Systems: Foundations and Principles*. New York, NY, USA: Kluwer, 2002.
[5] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*. New York, NY, USA: Oxford Univ. Press, 2017.
[6] R. Belohlavek and M. Krmelova, "Factor analysis of ordinal data via decomposition of matrices with grades," *Ann. Math. Artif. Intell.*, vol. 72, no. 1, pp. 23–44, 2014.
[7] R. Belohlavek and M. Trneckova, "The discrete basis problem and Asso algorithm for fuzzy attributes," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 7, pp. 1417–1427, Jul. 2019.
[8] V. V. Cross and T. A. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications*. Heidelberg, Germany: Physica-Verlag, 2002.
[9] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Inf. Control*, vol. 20, pp. 301–312, 1972.
[10] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York, NY, USA: Academic, 1980.
[11] D. Dubois and H. Prade, "Fuzzy cardinality and the modeling of imprecise quantification," *Fuzzy Sets Syst.*, vol. 16, pp. 199–230, 1985.
[12] F. Esteva, L. Godo, and C. Noguera, "A logical approach to fuzzy truth hedges," *Inf. Sci.*, vol. 232, pp. 366–385, 2013.
[13] J. A. Goguen, "The logic of inexact concepts," *Synthese*, vol. 18, pp. 325–373, 1969.
[14] P. Hájek, *Metamathematics of Fuzzy Logic*. Dordrecht, Netherlands: Kluwer, 1998.
[15] P. Hájek, "On very true," *Fuzzy Sets Syst.*, vol. 124, pp. 329–333, 2001.
[16] M. Holčapek, "Cardinalities of fuzzy sets and fuzzy quantifiers over residuated lattices," M.S. thesis, Dept. of Math., Faculty of Sci., Univ. of Ostrava, Moravská Ostrava a Přívoz, Czechia, 2005.
[17] A. Kaufmann, *Introduction à la Théorie des Sous-Ensembles Flous Tome 4: Compléments et Nouvelles Applications.* Paris, France: Elsevier Masson, 1977.
[18] G. Lakoff, "Hedges: A study in meaning criteria and the logic of fuzzy concepts," *J. Philos. Log.*, vol. 2, no. 4, pp. 458–508, 1973.
[19] G. L. Murphy, *The Big Book of Concepts*. Cambridge, MA, USA: MIT Press, 2004.
[20] D. Ralescu, "Cardinality, quantifiers, and the aggregation of fuzzy criteria," *Fuzzy Sets Syst.*, vol. 69, no. 3, pp. 355–365, 1995.
[21] V. Vychodil, "Truth-depressing hedges and BL-logic," *Fuzzy Sets Syst.*, vol. 157, no. 15, pp. 2074–2090, 2006.
[22] M. Wygralak, "An axiomatic approach to scalar cardinalities of fuzzy sets," *Fuzzy sets Syst.*, vol. 110, no. 2, pp. 175–179, 2000.
[23] M. Wygralak, *Cardinalities of Fuzzy Sets*. Berlin, Germany: Springer, 2003.
[24] L. A. Zadeh, "Fuzzy languages and their relation to human and machine intelligence," in *Proc. Int. Conf. Man, Comput.*, 1972, pp. 130–165.
[25] L. A. Zadeh, "A fuzzy-set-theoretic interpretation of linguistic hedges," *J. Cybern.*, vol. 2, no. 3, pp. 4–34, 1972.
[26] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning–I, II, and III," *Inf. Sci.*, vol. 8, pp. 199–249, 1975.
[27] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Sys., Man Cybern.*, vol. SMC-3, no. 1, pp. 28–44, Jan. 1973.
[28] L. A. Zadeh, "PRUF–A Meaning representation language for natural languages," *Int. J. Man-Mach. Stud.*, vol. 10, no. 4, pp. 395–460, 1978.
[29] L. A. Zadeh, "A theory of approximate reasoning," in *Machine Intelligence*, J. E. Hayes, D. Michie, and L. I. Mikulich, Eds. New York, NY, USA: Halstead Press, 1979, pp. 149–194.